DOCUMENT RESUME

ED 104 960 TM 004 431

AUTHOR McBride, James R.; Weiss, David J.
TITLE TETREST: A Fortran IV Program for Calculating Tetrachoric Correlations. Research Report 75-2.
INSTITUTION Minnesota Univ., Minneapolis. Dept. of Psychology.
SPONS AGENCY Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
REPORT NO RR-75-2
PUB DATE Mar 75
NOTE 49p.
AVAILABLE FROM Psychometric Methods Program, Dept. of Psychology, Univ. of Minnesota, Minneapolis, Minn. 55455 (while supplies last)

EDRS PRICE MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS *Computer Programs; *Correlation; Electronic Data Processing; Factor Analysis; Input Output; Item Analysis; *Matrices; Statistical Analysis; *Testing
IDENTIFIERS *Tetrachoric Correlation

ABSTRACT

A general purpose computer program for the calculation of a matrix of tetrachoric correlations is described. This program was developed for use in adaptive (and other) testing research for examining the unidimensionality assumption in latent trait theory, in conjunction with available factor analysis programs. Several other potential applications, as well as details for its use, are described. The program accepts as input raw dichotomous data, reduced joint frequency data, or joint and marginal proportions, for up to 75 items. Output options include the tetrachoric correlation matrix, the matrix of phi coefficients, fourfold frequency tables for every item pair, a joint frequency matrix (which reduces all the information in the fourfold tables to a square matrix with order equal to the number of items), and a pair-by-pair listing of input proportions and output correlations which permits testing the program against published tables of the tetrachoric correlation. Variable input and output formatting makes the program convenient to use in conjunctions with other analyses by packaged statistical programs. Examples of input and output are presented. A complete FORTRAN IV listing is included. (Author)

TM

TETREST: A FORTRAN IV PROGRAM FOR CALCULATING TETRACHORIC CORRELATIONS

James R. McBride

and

David J. Weiss

Research Report 75-2

Psychometric Methods Program
Department of Psychology
University of Minnesota
Minneapolis, MN 55455

March 1975

Prepared under contract No. N00014-67-A-0113-0029
NR No. 150-343, with the Personnel and
Training Research Programs, Psychological Sciences Division
Office of Naval Research

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 75-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TETREST: A Fortran IV Program for Calculating Tetrachoric Correlations		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) James R. McBride and David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-67-0113-0029
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, Minnesota 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE.:61153N PROJ:RR042-04 T.A.:RR042-04-01 W.U.:NR150-343
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE February 1975 13. NUMBER OF PAGES 39
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

tetrachoric correlation
bivariate normal correlation
fourfold point correlation
phi correlation
adaptive testing
test development
psychological testing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A general purpose computer program for the calculation of a matrix of tetrachoric correlations is described. This program was developed for use in adaptive (and other) testing research for examining the unidimensionality assumption in latent trait theory, in conjunction with available factor analysis programs. Several other potential applications, as well as details for its use, are described. The program accepts as input raw dichotomous data, reduced joint frequency data, or joint and marginal proportions, for up to 75

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) items. Output options include the tetrachoric correlation matrix, the matrix of phi coefficients, fourfold frequency tables for every item pair, a joint frequency matrix (which reduces all the information in the fourfold tables to a square matrix with order equal to the number of items), and a pair-by-pair listing of input proportions and output correlations which permits testing the program against published tables of the tetrachoric correlation. Variable input and output formatting makes the program convenient to use in conjunctions with other analyses by packaged statistical programs. Examples of input and output are presented. A complete FORTRAN IV listing is included.

Contents

Disclaimer

The computer program presented in this report has been tested carefully in the analysis of a variety of data. Although it has consistently yielded accurate results, the authors make no warranty as to its accuracy and functioning, nor shall the fact of its distribution imply such warranty.

TETREST: A FORTRAN IV PROGRAM FOR CALCULATING TETRACHORIC CORRELATIONS

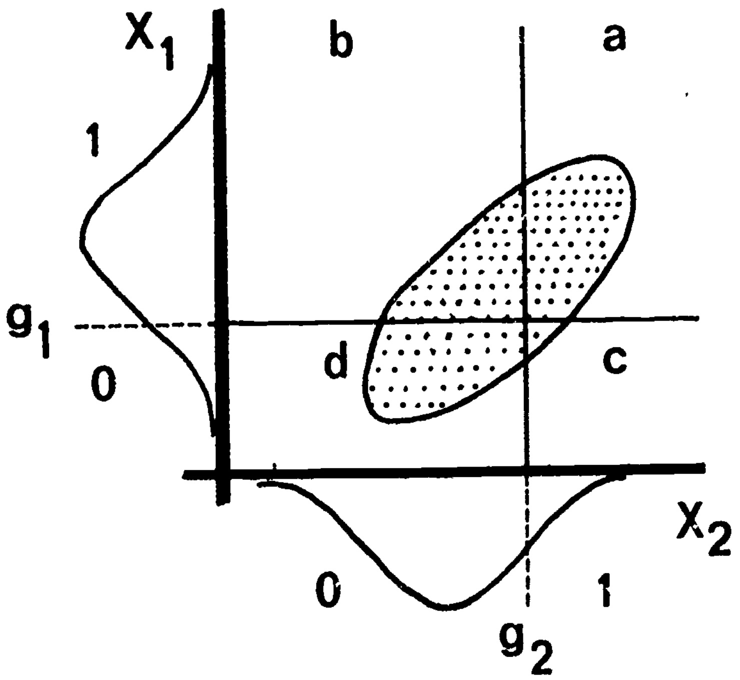
The Tetrachoric Correlation Coefficient

The tetrachoric correlation coefficient is an estimate of the product-moment correlation between two continuous normally distributed variables which have been dichotomized so that each variable takes only two discrete values. Stated another way, the tetrachoric r is an estimate from dichotomous data of the bivariate normal correlation between two continuous variables which have been dichotomized. Use of the tetrachoric correlation involves assumptions that 1) a continuous latent variable underlies the dichotomous data on each variable; 2) the latent variables are normally distributed; and 3) the regression of one on the other is linear.

These assumptions are illustrated in Figure 1, which depicts normally distributed variables X_1 and X_2 which have been dichotomized at points g_1 and g_2, respectively, in such a way that if the continuous value of X_n equals or exceeds g_n, a binary value of 1 is assigned; if X_n is less than g_n a value of 0 is assigned. Lines drawn perpendicular to the axes of X_1 and X_2, through

Figure 1

Partioning of a Bivariate Normal Distribution Scatterplot into Four Quadrants by Dichotomization



points g_1 and g_2, intersect to divide the bivariate normal scatterplot of

X_1 and X_2 into four quadrants labeled *a, b, c, d.*

The tetrachoric correlation is used to infer the bivariate normal correlation from the frequencies observed in each of the four quadrants. These frequencies are the joint frequencies which constitute a fourfold, or 2 x 2 contingency table of the two variables. Frequency *a* is the number of times that a value of 1 was observed on both variables jointly; *b* is the number of occurrences of a 1 on X_1 jointly with a 0 on X_2; *c* is the frequency of a (0,1) pattern on X_1 and X_2, respectively; *d* is the frequency of occurrence of the (0,0) pattern. Figure 2 illustrates the correspondence between the partitioned bivariate scatterplot and the fourfold table; each of the frequencies *a, b, c, d* corresponds to the observed frequency in the same-lettered quadrant of Figure 1.

Figure 2

The Fourfold (2x2) Table of Frequencies Corresponding to the Quadrants in Figure 1

		X_2		
		0	1	
X_1	1	b	a	a+b
	0	d	c	c+d
		b+d	a+c	

Figure 2 also depicts the marginal, or unconditional, frequencies of the two variables. For example, *a+b* is the unconditional frequency with which a value of 1 was observed on variable X_1; *a+c* is the unconditional frequency of a 1 on X_2. The marginal frequencies of a variable *n* are a function of the point of dichotomy, g_n. If the assumptions underlying it are met, the population tetrachoric correlation coefficient is invariant under changes in the point of dichotomy. This is not true of the phi coefficient, the only other correlation index used widely with dichotomous test items. The magnitude of phi is heavily dependent on the relative magnitudes of the marginal frequencies of the two variables.

The known lack of invariance of phi when there is diversity in the marginal frequencies from one item to another has led to increased use of the tetrachoric correlation. Whether tetrachoric correlation coefficients are invariant regardless of the points of dichotomy, when the assumption of bivariate normality is not met, is an empirical question whose answer is specific to the problem at hand. For bivariate normal data, the tetrachoric correlation estimates the population product-moment correlation and is invariant. For data which do not meet the assumption of bivariate normality the tetrachoric correlation can still be computed, but should be interpreted only as a measure of association whose invariance under disparate marginal frequencies is questionable.

Phi coefficients and tetrachoric correlations also differ in another important respect. A matrix of sample phi coefficients is always non-negative definite, or Gramian. This means that such a matrix is appropriately structured for the application of certain statistical analysis techniques, including factor analysis. Matrices of sample tetrachorics are often non-Gramian (Lord and Novick, 1968, p. 349). Factor analysis of a non-Gramian correlation matrix of full rank will result in one or more factors having negative eigenvalues; that is, "factors" which account for negative variance. Related problems may include communality estimates exceeding unity, factor loadings exceeding unity, and inability to estimate communalities iteratively because the inverse of the correlation matrix is indeterminant.

The non-Gramian property of some sample tetrachoric correlation matrices may be due to violating the normality assumptions for one or more variables, to sampling error where the population interitem distributions are all bivariate normal, or to numerical errors in estimating the correlations. Any or all of these sources of error may render hazardous the interpretation of factor analyses based on tetrachoric correlation matrices. Users of the tetrachoric correlation should be aware of these potential problems, and should exercise caution in statistical analyses of tetrachoric correlation matrices.

Applications of the Tetrachoric Correlation Coefficient

Most recent research in adaptive testing has employed a latent trait theoretical model to account for testees' responses to dichotomous test items. The most prevalent latent trait models assume that the latent variable underlying test item responses is one-dimensional. For constructing test item pools, this assumption requires that the latent space spanned by the items be unidimensional. A sufficient condition for the unidimensionality of the latent space is that the matrix of tetrachoric item intercorrelations has just one common factor (Lord and Novick, 1968, pp. 381-382). For practical purposes, the items are considered to represent a unidimensional variable if the largest latent root of the sample matrix of their tetrachoric intercorrelations accounts for a large proportion of the common variance, and the second and smaller latent roots are of about the magnitude that might be expected from sampling errors (Indow & Samejima, 1966; McBride & Weiss, 1974).

The tetrachoric correlation, then, is a useful statistic in the development and analysis of item pools for adaptive testing. But it is useful for other applications where dichotomously scored items are used, by virtue of its invariance properties. If the assumptions underlying it are met, the tetrachoric correlation coefficient is invariant under differences in the marginal proportions of the two variables involved. As discussed above, this invariance property is not the case with the fourfold point (phi) correlation coefficient, which reaches its maximum value only if both variables of a pair have the same marginal proportions (Carroll, 1961). For this reason, the tetrachoric correlation is useful in factor analyses of dichotomous variables whose marginal proportions vary widely.

Another application of the tetrachoric *r* is in the estimation of the parameters of the item characteristic curve in latent trait test theory. If all the items in a test are considered to have item characteristic curves of normal ogive form (Lord & Novick, 1968, Chapter 16), Bock and Wood (1971) have pointed out that under certain conditions each item's discrimination parameter may be estimated from its loading on the first factor extracted from the matrix of tetrachoric interitem correlations.

Despite the increase in applications of the tetrachoric correlation, general purpose computer programs for calculating it are not widely available. The computer program described in this report was written specifically to fill this void. Its original application permitted a unidimensionality analysis of a set of test items being assembled in order to implement an adaptive ability testing program (McBride & Weiss, 1974). It is presented here in order to make a tetrachoric correlation program available to users who would otherwise not have access to one.

General Description of TETREST

This program (TETRachoric ESTimation) uses an approximation procedure given by Kirk (1973) to estimate the tetrachoric correlation coefficient between two dichotomized or dichotomous variables. It also calculates the value of the fourfold point (phi) correlation between two dichotomous variables.

TETREST is designed primarily to:

1. Construct a set of four fourfold (2 x 2) contingency tables, for every pair of variables, from dichotomous data on a set of K variables. For K variables there are $[(K)(K-1)]/2$ such variable pairs.

2. Estimate the degree of relationship (tetrachoric and/or phi correlation) in each fourfold table.

3. Construct the KxK matrix of tetrachoric and/or phi correlation coefficients in a form amenable to factor analysis using packaged computer statistical routines such as SPSS: Statistical Package for the Social Sciences (Nie, Bent and Hull, 1970) or BMD: Biomedical Computer Programs (Dixon, 1973).

The program is general, emphasizing flexibility and ease of input and output for the user. It was written specifically for the CDC CYBER 74 computer in Control Data FORTRAN IV. TETREST will accept input data from punched cards, magnetic tape, or disc storage. It will print, punch, or write its output on logical units in several different forms. The specific input and output options are described below.

Data Input

TETREST will accept input data of three different types, all arising from dichotomous data on K variables observed over N individuals. The three data types are illustrated in Figures 3, 4, and 5 respectively.

Type I is the basic form of data: the dichotomous scores of N individuals on K items. Figure 3a illustrates the Type I data of 10 individuals on three

Figure 3

(a). Type I data from a 3-item test for 10 individuals

Individual	Item 1	Item 2	Item 3
1	1	0	0
2	1	1	0
3	1	1	0
4	0	0	1
5	1	1	0
6	0	0	1
7	1	1	0
8	1	0	1
9	1	0	1
10	0	0	1
Totals	7	4	5

(b). The fourfold (2x2) contingency tables for each pair of variables, derived from Type I data

		Item 2: 0	Item 2: 1	Total
Item 1	1	3	4	7
	0	3	0	3
Total		6	4	10

		Item 3: 0	Item 3: 1	Total
Item 1	1	5	2	7
	0	0	3	3
Total		5	5	10

		Item 3: 0	Item 3: 1	Total
Item 2	1	4	0	4
	0	1	5	6
Total		5	5	10

dichotomous items (variables). Each individual's item scores (1 or 0) are on a separate data record, such as a punched data card. Counting the frequency of 1's and 0's permits construction of the fourfold tables, shown in Figure 3b.

Type II data is the reduction of Type I data to a KxK matrix of frequency counts. Figure 4 shows the Type I data of Figure 3 reduced to a 3 x 3 matrix, called the joint frequency matrix, which contains the minimum frequency information needed to construct the fourfold tables for each of the $[(K)(K-1)]/2$ possible pairs of variables. The three diagonal entries (7,4,5) are simply the three marginal "positive" frequencies (e.g., the number of correct or endorsed items). The supra-diagonal entries (4,2,0) are the joint positive (i.e., 1,1) frequencies for variable pairs (1,2) (1,3) and (2,3) respectively; that is, the frequency in cell a of the fourfold table of variables i and j.

The infra-diagonal entries (3,0,1) are the joint negative frequencies (i.e., 0,0) for variable pairs (2,1), (3,1) and (3,2) respectively; that is, the frequency in cell *d* of the fourfold tables (see Figure 2).

Figure 4

The Joint Frequency Matrix of Type II Data

		1	2	3
	1	7	4	2
Item	2	3	4	0
	3	0	1	5

Figure 5 shows Type III data reduced from the fourfold tables in Figure 3b. Type III data consists of three proportions (P_{ij}, q_i, q_j) and and optional sign-change character, SIGN. Each set of three proportions is derived from the fourfold table of one variable pair.

P_{ij} is the joint proportion obtained by dividing one of the cell frequencies (*a*,*b*,*c* or *d*) by *N*, the number of individuals observed ($N=a+b+c+d$). The appropriate cell is determined by the marginal proportions. P_{ij} is the joint proportion occurring in the cell corresponding to the smaller marginals.

q_i is the smaller marginal proportion for variable i.

q_j is the smaller marginal proportion for variable j.

SIGN is a character which changes the sign of the obtained correlation. This is necessary whenever P_{ij} comes from cell *b* or *c* of the fourfold table. If SIGN is left blank, no sign change will occur. Any other character, such as a minus sign (-), will effect the sign correction.

Figure 5

The Joint Proportions Matrix of Type III Data

Variable Pair (i,j)	P_{ij}	q_i	q_j	SIGN
(2,1)	0	.40	.30	-
(3,1)	.30	.50	.30	-
(3,2)	0	.50	.40	

Data of Type I will likely be used most often in calculating tetrachoric correlation coefficients. TETREST is capable of yielding data of Type II as output since some program users might desire to use such output as input at a later time. Where data of Type II are available, their use as input can represent a considerable saving of computer processing time, especially when the number of individuals in the original data is very large. The provision for data of Type III permits rapid calculation of tetrachoric or phi correlations for one or more variable pairs for which raw data or frequency matrix data are not available. One application of Type III data might be to check TETREST results against tables of the tetrachoric correlation coefficient, such as those published by the National Bureau of Standards (1959).

Data Output

TETREST is capable of several output options; the output ε ι depends in part on the type of input data used.

For Type I and Type II data, the following output options are available:

1. Printing or writing on tape or disc of the *KxK* joint frequency matrix (i.e., Type II input).

2. Printing of the $[(K)(K-1)]/2$ 2 x 2 frequency tables, including values of phi and tetrachoric *r*.

3. Printing, punching, or writing on tape or disc or the *KxK* matrix of tetrachoric and/or phi intercorrelations in either lower diagonal or square format.

For Type III data, option 3 is available. Also available optionally for Type III data is a listing by variable pair of the input proportions, tetrachoric and phi correlation values, and the degree of convergence occurring in the approximation of the tetrachoric *r*.

Program Use

TETREST incorporates a number of input and output options which may be specified by the user. Implementing these options involves the use of from five to ten program control cards. Each card has a task definition field consisting of the first 15 columns of the card, and one or more specification fields spanning the remaining 65 columns (16-80). Program users familiar with the SPSS system (Nie, Bent, and Hull, 1970) will recognize that the present system of control cards is very similar to the one used within SPSS. The task definition field of each control card defines the nature of the data appearing in the specification fields. The ten program control cards, in the order in which they should appear, are as follows:

PROBLEM	(mandatory)
RUN NAME	(optional)
INPUT MEDIUM	(optional)
INPUT FORMAT	(mandatory)
NO. OF CASES	(optional)
OUTPUT FORMAT	(optional)
OPTIONS	(optional)
READ INPUT DATA	(optional)
MATRIX	(mandatory)
FINISH	(optional)

Descriptions of the control functions of each card appear in sections A through J below.

A. The PROBLEM Card (mandatory)

This card may contain from 1 to 6 specification fields, as shown below:

```
PROBLEM         NVAR=--   DATA=-    2X2=-    NFMT=--    KEY=-      R=----
```

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80

The data in each field of the PROBLEM card must appear in exactly the columns specified.

NVAR=_ _specifies the number of variables in the analysis. NVAR must be a right-justified integer number greater than 1 and less than 76. NVAR is the only mandatory specification on the PROBLEM card.

DATA=_ specifies the type of input data and some output information. If DATA is not specified, Type I data is assumed. DATA may take any integer value from 1 to 5, or may be left blank. These specification values result in the following input/output options:

DATA=1 or blank: the program will read Type I data.
- 2 : the program will read Type I data and output the joint frequency matrix calculated.
- 3 : the program will read Type II data.
- 4 : the program will read Type II data and output the joint frequency matrix.
- 5 : the program will read Type III data.

2x2=9 causes the program to print the [*(K)(K-1)*]/*2* 2x2 frequency tables, one for every variable pair, only if Type I or Type II data was specified. Any other numerical value, or a blank, suppresses printing of the 2x2 tables. Included in each 2x2 table are the intervariable phi and tetrachoric correlation coefficients. (CAUTION: each 2x2 table requires 6 lines of print. A 75-variable problem will print [(75)(74)]/2=2775 tables, which requires over 200 pages of print for the 2x2 tables alone.)

NFMT=_ _specifies the number of cards used to specify the input format. NFMT may be a right-justified integer number from 1 to 10. If NFMT is not specified, it is set equal to 01, and the variable input format must be no longer than 65 characters. (See INPUT FORMAT below.)

KEY=_specifies the data character which will be considered the "correct" or "positive" value in the analysis of Type I data. For example, in scored test data the character *1* might be assigned a correct response; in questionnaire data a *Y* might represent a "yes" response. KEY may be any alphanumeric character. Once KEY is specified, any input data character other than KEY will be considered as "incorrect" or "negative." For example, in *1-0* (binary) data a *1* would be considered correct if so specified; and a *0*, blank, or any other character would be considered incorrect.

If KEY is not specified, it is automatically set to *1* (a Hollerith constant). If data other than Type I data are used, KEY is ignored.

R=_ _ _ _specifies the correlation coefficient(s) to be used in constructing the output correlation matrix or matrices. Either tetrachoric or phi correlation matrices, or both, may be specified by a left-justified alphabetical entry to the right of R=.

R=PHI specifies that only a matrix of phi coefficients will be calculated.

R=TET or a blank field specifies that only a matrix of tetrachoric correlations will be calculated.

R=BOTH specifies that both matrices will be calculated.

B. The RUN NAME Card (optional)

This card allows the program user to print any arbitrary 65-character label at the top of the first printed page of the data analysis. The label must be punched in columns 16-80 of the RUN NAME card. The RUN NAME card, which takes the form shown below, is completely optional.

```
RUN NAME       ANY ARBITRARY SET OF CHARACTERS
```

C. The INPUT MEDIUM Card (optional)

This card allows the program user to input data from cards, magnetic tape, or disc storage file. The specification must begin in column 16 of the card, and must be one of the three values shown below in the example. If an INPUT MEDIUM card is not used, the program will read the input data from the INPUT file, which ordinarily implies punched cards.

```
INPUT MEDIUM   CARD    read data from file INPUT
               DISC  } read data from file DATA
               TAPE  }
```

If INPUT MEDIUM specifies DISC or TAPE, the user must supply the input data on a file named DATA, and must position that file at the start of the first record of the data to be read. (The program does not rewind the file DATA. This is to permit separate analyses of sequential blocks of data on the same file if desired by user.)

D. The INPUT FORMAT Card (mandatory)

The program reads data in an input format specified on this card by the user. The program uses a variable format, the form of which varies with the type of data to be analyzed. The INPUT FORMAT card is mandatory for the first problem to be analyzed, but is optional thereafter unless the format changes.

Type I data must be read in A1 format fields, since each variable score will be compared with the one-character alphanumeric value of the KEY specification.

Type II and III data (except SIGN) must be read in floating-point numeric fields (F-fields), since the data are numeric and will be read into floating-point arrays. SIGN must be read in an A1 format.

The INPUT FORMAT specification begins in column 16 with a left parenthesis, (, and the input format, which begins on the INPUT FORMAT card and continues through NFMT cards, as specified on the PROBLEM card. The input format may extend to as many as 10 cards in length, inclusive of the INPUT FORMAT card. Continuation cards of the INPUT FORMAT may use the entire 80 card columns. The format specification must end with a right parenthesis,).

```
INPUT FORMAT   (..........)
```

E. The NO. OF CASES Card (optional)

This card specifies the number of individual records of NVAR variable scores for the program to read in using the INPUT FORMAT specified by the user, in the case of Type I data only. The specification field begins in column 16 of the control card, and must be a right-justified integer number. E.g.,

```
NO. OF CASES   0192
```

The program interprets this as 192 subjects. The program will read exactly the number of cases of Type I data specified, unless it first encounters an end-of-file on the input data file. In that case it resets the number of cases to the number of individual data sets read before the end-of-file was encountered. A message to this effect appears on the control-card page if an end-of-file is encountered. Program users may employ this feature to their advantage when estimating the number of cases if the exact number is not known, simply by using any reasonable over-estimate of the number of cases.

If a NO. OF CASES card is not read, the number is automatically set to 9999, and the program will read 9999 individual data records, or until it finds an end-of-file mark as described above. CAUTION: In Type I data, the number of cases is reset to 9999 at the end of each problem, and must be respecified using the control card in each succeeding problem when appropriate. The number of cases has no bearing on the solution of Type III data problems, since data are input as proportions.

F. The OUTPUT FORMAT Card (optional)

The format in which the matrices of tetrachoric and/or phi correlation coefficients will be printed and/or written on the logical output unit may be specified using this card. The card has the form:

```
OUTPUT FORMAT  (..........)
```

The specified format must be enclosed in parentheses, and must be contained entirely in columns 16 through 80 on the OUTPUT FORMAT card. The format should consist of F-fields, in order to provide for the floating-point numerical values of the correlation coefficients.

If no OUTPUT FORMAT card is read, the program will write and/or print matrices in 8F10.7 format.

G. The OPTIONS Card (optional)

The specification on the OPTIONS card determines the disposition of the correlation matrices resulting from the program's calculations. The value specified must be an integer number between 1 and 6, inclusive. These values cause the following dispositions of the r-matrices:

OPTIONS	1	PUNCH the r-matrix on data cards
	2	PUNCH and PRINT the r-matrix
	3	PRINT the r-matrix on the output file
	4	PRINT the r-matrix and WRITE it on TAPE8
	5	WRITE the r-matrix on file TAPE8
	6	suppress any output of the r-matrix

Defult value of the OPTIONS card is 3; i.e., printed output only.

NOTE: 1. The OUTPUT FORMAT, whether set by the program or specified, applies only to printing or writing the r-matrix. If PUNCH is specified (option 1 or 2), the PUNCH format is 8F10.7 whenever square matrix output is chosen (in keeping with the SPSS factor analysis program's requirements for a square correlation matrix in that format). When the lower diagonal matrix is selected (via the STATISTICS card), punched output is in format (10F8.5), which is somewhat more economical with little loss of accuracy. See section I, below, for information on the square and lower diagonal matrix output options.

2. For options 4 and 5, each correlation matrix is written as a file on TAPE8, separated from other files by an end-of-file mark. When both tetrachoric and phi matrices are specified (via R=BOTH on the PROBLEM card) the tetrachoric matrix is written first, the phi matrix second.

H. The READ INPUT DATA Card (optional)

This card has no specification field, and has the following form:

READ INPUT DATA

Its function is twofold:

1. It causes the program's internal data file (TAPE1) to be rewound (i.e., repositioned at the start of the first logical record), effectively erasing the input data from the previous problem, if any.

2. It causes the program to read from the user's data file (INPUT or DATA) the input data of the type specified on the PROBLEM card.

For Type I data, the READ INPUT DATA card causes the program to read N (=NO. OF CASES) individual records of K (=NVAR) variable scores from the input data source specified, according to the INPUT FORMAT specified. If an end-of-file mark is read before the <u>*Nth*</u> data record, N is reset as described in section E, above.

For Type II data, the READ INPUT DATA card causes the program to read K (=NVAR) records of K frequencies each (i.e., a KxK joint frequency matrix) according to NVAR as specified on the PROBLEM card, and in the INPUT FORMAT specified. If an end-of-file mark is encountered before the *Kth* full record is read, processing is terminated, and an error diagnostic is printed.

For Type III data, the READ INPUT DATA card causes the program to read $[(K)(K-1)]/2$ separate sets of proportions, one set for each pair of variables, and later to process them as though they were infra-diagonal entries in a KxK matrix of proportion sets. Thus the first set results in the obtained correlation being treated as though it were between variables 2 and 1; the second between 3 and 1; the third between 3 and 2; the fourth between 4 and 1, and so on. The example below is for a 4-variable problem with [(4)(3)]/2=6 variable pairs.

	Data	Variable Pair
INPUT data set 1	$p, q_1, q_2,$ sign	(2,1)
2	$p, q_1, q_2,$ sign	(3,1)
3	$p, q_1, q_2,$ sign	(3,2)
4	$p, q_1, q_2,$ sign	(4,1)
5	$p, q_1, q_2,$ sign	(4,2)
6	$p, q_1, q_2,$ sign	(4,3)

r-matrix:

variable	1	2	3	4
2	1			
variable 3	2	3		
4	4	5	6	

The numbers within the r-matrix schematic above correspond to the input data sets of proportions (see Figure 5).

If no r-matrix output is desired (i.e., OPTION 6 is chosen), the order of input of the entries need not follow the above example. If the r-matrix is to be output, however, the program user must be careful of the order of his input data proportion sets, since their position in the r-matrix is determined entirely by their order in the input data.

If an end-of-file mark is encountered before a complete matrix of $[(K)(K-1)]/2$ Type III data sets is read, K is arbitrarily truncated to one more than the smallest value for which a square data matrix exists in the data. This will permit processing of all the proportional data, but will result in the use of indefinite operands if an attempt is made to output the r-matrix. Therefore, if proportional data (Type III) is used as input and the r-matrix is required as output, the user must be sure he includes exactly $[(K)(K-1)]/2$ sets of proportions, and incorporates them into his data in the correct order. If the r-matrix output is not required, any overestimate of NVAR=K will suffice.

Omitting the READ INPUT DATA card results in the specified analyses being performed on whatever data (if any) are stored on file TAPE1. For the initial problem in a set of one or more problems, TAPE1 is an empty file until the READ INPUT DATA card is read. (A "problem" is defined as the analysis specified in the set of control cards between the PROBLEM card and the MATRIX card inclusive. More than one "problem" may be included in a single computer run.) After the initial problem, however, TAPE1 contains whatever data was input for the preceding problem via the INPUT FORMAT card. Therefore, specifying a "problem" without including the READ INPUT DATA card, results in a second pass through the input data of the preceding problem. By varying NVAR, NO. OF CASES, and the INPUT FORMAT the program user can perform analyses of any subset of the initial problem's data, if the READ INPUT DATA card is omitted.

I. The MATRIX Card (mandatory)

Three options may be specified using the MATRIX card. The specification must be an integer number between 1 and 3, in the following format.

MATRIX 1 — For Type III data, print the input proportions, tetrachoric and phi correlations, one variable pair at a time. Output the r-matrix in lower diagonal form.

2 — Output r-matrix in lower diagonal form only. (Suppress Individual variable pair results if Type III data are involved.)

3 — Output r-matrix in square form. (Suppress individual variable pair results if Type III data are involved.)

The MATRIX card must be the last card in a problem set (excepting the FINISH card), since execution of the data processing and output options is begun after the MATRIX card is read. Placing the MATRIX card out of place will result in all cards following it being read as control cards for the succeeding problem or as data cards. Placing it before the READ INPUT DATA card will result in the analyses specified in the present problem being performed on the data of the preceding problem, or on an empty data file, as described above.

If no matrix output is desired (i.e., OPTION 6 is elected) any MATRIX option (1,2 or 3) may be used to signal the start of data processing.

J. The FINISH Card (optional)

This card has no specification field, and takes the form:

FINISH

When the FINISH card is encountered on the INPUT file, processing terminates, and control is returned to the computer operating system. Thus the FINISH card should appear only as the last card in a set of program control cards; it may not be used to separate problems. Failure to include a FINISH card results in printing an error diagnostic, program termination, and return of control to the computer operating system. In general, this will not affect the outcome of the user's programs other than TETREST, so the FINISH card may be considered optional. Its use is recommended whenever any data follow the TETREST control cards on the INPUT file.

K. Control Card Order

The following are the only restrictions on control card order within a problem. (A problem is the execution of the program on a single set or matrix of data. Multiple problems may be included in a single run of the program.) Aside from these restrictions, control cards may be placed in any order.

1. The READ INPUT DATA card, if present, must follow the INPUT MEDIUM, INPUT FORMAT and NO. OF CASES cards whenever any or all are used.

2. Data cards, if used, must follow the READ INPUT DATA card. If the input medium is CARD, and if the program is to count observations (number of cases), then an end-of-record card (7-8-9) must be inserted between the last data card and the next program control card. Failure to comply with this requirement will result in program control cards being read as data cards, and will usually terminate program execution without a diagnostic.

   On the other hand, if the number of cases is specified exactly, no end-of-record card should separate the data cards from the control cards.

3. The MATRIX card must be the last control card in each problem.

4. The FINISH card must be the last card in a run.

L. Examples of Control Card Use

1. Card input of Type I data; program counts number of cases; a lower diagonal tetrachoric correlation matrix will be output.

```
RUN NAME        EXAMPLE 1
PROBLEM         NVAR=05                 R=TET
INPUT FORMAT    (5A1)
OPTIONS         3
READ INPUT DATA
10101
11100
.....
.....
00111
>               (RECORD SEPARATOR)
MATRIX          2
```

2. Card input of Type I data; number of cases specified exactly by user; a square phi correlation matrix will be output.

```
RUN NAME        EXAMPLE 2
PROBLEM         NVAR=05                R=PHI
INPUT FORMAT    (5A1)
NO. OF CASES    0005
OPTIONS         3
READ INPUT DATA
10101
11100
10111
01100
10110
MATRIX          3
```

3. Tape input of data; two problems in succession (one on cases 1-100 in the data file; the second on cases 101-200); both tetrachoric and phi matrices will be output in lower diagonal format.

```
RUN NAME        FIRST 100 CASES
PROBLEM         NVAR=05                R=BOTH
INPUT MEDIUM    TAPE
INPUT FORMAT    FIXED
INPUT FORMAT    (5A1)
NO. OF CASES    0100
OPTIONS         2
READ INPUT DATA
MATRIX          2
RUN NAME        SECOND 100 CASES
PROBLEM         NVAR=05                R=BOTH
INPUT MEDIUM    TAPE
NO. OF CASES    0100
OPTIONS         2
READ INPUT DATA
MATRIX          2
FI  SH
```

4. The above three examples are merely suggestive of the flexibility of options available to the program user. Far too many different configurations of control cards are permissible to list them all here. The following example uses the minimum number of control cards, setting program parameters by default everywhere possible:

```
PROBLEM         NVAR=05
INPUT FORMAT    (5X,5A1)
READ INPUT DATA
                INSERT DATA HERE
>               (RECORD SEPARATOR)
MATRIX          2
FINISH
```

This minimal configuration assumes Type I input data punched on cards, with the program counting number of cases. It will print the tetrachoric correlation matrix only, in lower diagonal form and 8F10.7 format, on the file OUTPUT.

Additional examples of program input, in conjunction with the output they generated, are shown in Appendix A.

Program Specifications

File Structure

Eleven files are used within TETREST, in the following order:

INPUT

OUTPUT

PUNCH

DATA	the user-supplied file containing the data to be analyzed whenever the INPUT MEDIUM is DISC or TAPE.
TAPE1	the file used within the program to store the user's data. When the READ INPUT DATA control card is read, TAPE1 is rewound, and data read from file INPUT or DATA is written onto file TAPE1.
TAPE8	the file on which the r-matrix is written when the user chooses OPTIONS 4 or 5. TAPE 8 is never rewound within the program; an end-of-file is written on it after each matrix is written on it. The user must rewind TAPE8 in order to have access to the data on it.
TAPE9	the file on which the joint frequency matrix is written when the user specifies DATA=2 or 4 on the PROBLEM card. The comments above for TAPE8 apply to TAPE9 as well.
TAPE6=OUTPUT	TAPE6 is a logical unit, set equivalent to the file OUTPUT.
TAPE10=PUNCH	TAPE10 is a logical unit, set equivalent to the file PUNCH.
TAPE51=DATA	TAPE51 is a logical unit, set equivalent to the file DATA.
TAPE55=INPUT	TAPE55 is a logical unit, set equivalent to the file INPUT.

Subprograms

TETREST includes the following subroutines and functions as part of the program:

TETREST	The main program; executes the input and output options.

SETUP	Reads the program control cards and the input data; determines the desired input/output options; issues error diagnostics.
FREQ1	Constructs the *KxK* joint frequency matrix when Type I data are used.
CELLS2	Constructs a 2x2 contingency table for each pair of variables; selects the appropriate frequencies for the analysis; constructs the r-matrix.
APXTET3	Approximates tetrachoric *r* by means of iteration given the joint and marginal proportions.
GAUSS8	(Function) Performs 8-point Gaussian quadrature.
MTRXOUT	Outputs the appropriate correlation matrices in square or lower diagonal form.

The following system library subroutines are used within TETREST:

DATE	subroutine returns the current day, month and year.
CDFNI	cumulative normal distribution function inverse.
EXP	the exponential function.
ABS	absolute value function.
SQRT	the square foot function.

Additionally, TETREST makes use of the Control Data ENCODE and DECODE statements, as well as the IF(EOF) end-of-file check.

Timing and Core Memory Requirements

TETREST requires 46000 octal core memory (CM) cells to run when using the FORTRAN extented (FTN) compiler under KRONOS 2.1 on the CDC CYBER 74 computer.

Central processor (CP) time for a TETREST run is dependent on the data type, the correlation coefficient chosen, and the number of variables.

Type I data requires the most CP time; Type III requires the least. For Type I data, CP time varies directly with the NO. OF CASES and with NVAR, the number of variables. CP time for the other data types is a direct function of NVAR only.

No formula is available for forecasting CP time requirements. Actual data runs have required CP time as listed below:

Case	Data Type	NVAR	NO. OF CASES	CP Seconds
1	I	75	80	15
2	I	60	200	15
3	I	50	240	25
4	I	20	200	7

The disparity in CP time between Case 3 and Cases 1 and 2 illustrates the difficulty of prescribing a rule of thumb. CP time also depends on idiosyncracies in the data which affect the average number of iterations required per correlation.

Availability

A complete FORTRAN listing of the program is included in Appendix B. Card decks are available from the authors. A copy of the source code will be provided on a tape submitted by the prospective user.

Method of Calculation

Tetrachoric Correlation Coefficient

The tetrachoric correlation between two dichotomous variables involves solution of the following equation for r, given the values of L, h,k,x, *and* y:

$$L(x,y,r)=\int_x^\infty\int_y^\infty\left(2\pi\sqrt{1-r^2}\right)^{-1}\exp\left[\frac{-(h^2+k^2-2rhk)}{2(1-r^2)}\right]dh\,dk \qquad (1)$$

where r is the tetrachoric correlation

x,y are standard normal deviates

L is a joint function of the two dichotomous variables and their tetrachoric relationship (whose numerical value is observable as a joint proportion)

The values of x and y, h and k are obtained from calculations based on the marginal proportions of the two dichotomous variables, which can be obtained in turn from their 2x2 contingency table:

		Variable j: incorrect (0)	Variable j: correct (1)	total	proportion
Variable i	Correct (1)	b	a	a+b	p_i
	Incorrect (0)	d	c	c+d	$(1-p_i)$
		b+d	a+c		
	proportion	$(1-p_j)$	p_j		

The parameters of Equation 1 are estimated from the sample contingency table as follows:

$x=\Phi(q_i)$=the standard normal deviate corresponding to $q_i (=1-p_i)$

$y=\Phi(q_j)$=the standard normal deviate corresponding to $q_j (=1-p_j)$

L=d/N=the joint proportion of persons responding positively to both variables (items)

In general, approaches to solving Equation 1 have involved the use of an infinite series. McNemar (1955) gives one such solution, expressed here in notation slightly different from his:

$$\frac{\frac{t}{N} - q_i q_j}{hk} = r + xy\frac{r^2}{2!} + (x^2-1)(y^2-1)\frac{r^3}{3!} + \cdots \qquad (2)$$

where h,k are the ordinates of the standard normal curve at x and y, respectively

McNemar's formulation includes the restriction that q_i and q_j (the marginal proportions) must each be less than or equal to 1/2. Obviously, that restriction implies that t may be the joint frequency value occurring in cell a, b, c, or d of the 2x2 table, whichever cell corresponds to the two marginal values meeting the restriction. Use of an inappropriate cell to obtain the joint frequency t will introduce an error in the proportion t/N and hence will in general result in an incorrect numerical solution for r. TETREST, in dealing with Type I data, constructs the 2x2 table for each pair of variables and selects the appropriate joint frequency for correct solution of Equation 2.

Kirk (1971, 1973) devised an approximation procedure based on Gaussian quadrature and use of a Newton-Raphson iteration procedure to evaluate the integral and approximate r. The first two terms of Equation 2 are used to obtain an initial estimate based on the empirical values of t/N, q_i and q_j.

In order to employ Gaussian quadrature and restrict the range of the estimate of r, Kirk gives the following transformation, based on Equation 1.

$$f(r) = \frac{r}{2\pi}\int_0^1 \frac{1}{\sqrt{1-u^2r^2}} \exp\left[\frac{-(x^2-2xyur+y^2)}{2(1-u^2r^2)}\right] du \qquad (3)$$

$$\doteq \frac{r}{2\pi} \sum_{i=0}^{n} w_i g(u_i) \qquad (4)$$

$$\text{where } g(u) = \frac{1}{\sqrt{1-u^2r^2}} \exp\left[\frac{-(x^2-2xyur+y^2)}{2(1-u^2r^2)}\right]$$

and the u_i are the roots of the Legendre polynomials; the w_i are the associated weights for an (n+1)-point quadrature.

The starting value for solution of Equation 4 is obtained by direct solution of the following equation for r_0:

$$\frac{P-q_1q_2}{hk} = r_0+xy\,\frac{r_0^2}{2!} \tag{5}$$

where P=t/N, taken from the appropriate cell of the 2x2 table. A limiting value of $\pm.80$ is used whenever the absolute value of the starting estimate exceeds 1.00. Successive values are calculated using Newton-Raphson iteration:

$$r_{i+1} = r_i - \frac{f(r_i)-m}{f'(r_i)} \tag{6}$$

where $m=P-q_1q_2$

and

$$f'(r) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left[\frac{-(x^2-2xyr+y^2)}{2(1-r^2)}\right] \tag{7}$$

Iteration continues through as many as 100 cycles or until

$$\left|r_{i+1}-r_i\right| < .0001 \tag{8}$$

Convergence failures can occur. When convergence to the criterion of .0001 does not occur, one of two alternatives takes place:

1. If convergence to within .001 has occurred, an estimate of r is returned, but an informative diagnostic is printed, telling the user the indices of the two variables involved and the degree of convergence attained.

2. If convergence to within .001 has not occurred, a value of ±99.00 is returned for r.

Generally, failures of convergence are due to extremely low values of the joint proportion, P, or to the joint proportion P being very close to either of the marginal proportions, q_i or q_j. One circumstance in which this is a problem is the case in which one or more cells in the 2x2 table contains a zero.

When three zero cells occur, as in the following

		j		
		1	0	
i	1	a	0	p_i=1
	0	0	0	q_i
		p_j=1	q_j	

the tetrachoric correlation is automatically set to zero, since there is no covariance. Similarly, when there is zero variance in only one variable (p_i=1.00 or p_j=1.00) there is zero covariance between variables, and the tetrachoric correlation is zero.

When just two zero cells occur, and these occupy either diagonal of the 2x2 table, the program returns a value of $\pm$1.00, the sign being determined by the sign of the covariance between the two variables. For example:

		j		
		1	0	
i	1	0	b	p_i
	0	c	0	q_i
		p_j	q_j	

here the covariance equals $0-p_ip_j$, and the correlation coefficient is returned as -1.0.

In the case of a single zero cell, such as

		j	
		1	0
i	1	0	b
	0	c	d

there is no rational basis for setting the tetrachoric r to any fixed value as was done above. There is non-zero covariance, as well as variance in both variables. This class of problems frequently results in failure of the iteration procedure to converge on a value of the tetrachoric r. Non-convergence is especially a problem with a single zero cell concurrent with an extreme proportion in both variables. As indicated earlier, a failure of convergence is signalled by a value of $\pm$99.0 being returned in place of the tetrachoric r.

Kirk's algorithm for approximating the value of the tetrachoric correlation does not directly calculate the correct sign of the relationship. For Type I and Type II data, the sign of the tetrachoric correlation is determined by the sign of the covariance, which is completely determined by the determinant *ad-bc* of the fourfold table of any variable pair. Thus, when *ad-bc* is positive, a positive value for the tetrachoric is output; a negative value is output when *ad-bc* is negative.

For Type III data the sign of the tetrachoric *r* is corrected by the sign-change character, SIGN. If SIGN was blank on input, the value calculated by TETREST is returned uncorrected. If SIGN was any non-blank character, the calculated value of the tetrachoric *r* is multipled by -1 to correct the sign of the relationship.

The Phi Coefficient

For Type I and II data, the value of the phi coefficient is calculated from the frequencies of the fourfold table, using the following formula:

$$PHI = \frac{ad-bc}{\sqrt{(a+b)(a+c)(b+c)(b+d)}} \qquad (9)$$

In the case of Type III data, phi is calculated from the input proportions and SIGN:

$$PHI = \frac{p_{ij} - q_i q_j}{\sqrt{(q_i)(1-q_i)(q_j)(1-q_j)}} \qquad (10)$$

where SIGN = +1 or -1 according to the input convention described above.

Accuracy

The method used for estimating the tetrachoric correlation coefficients in TETREST is a numerical approximation technique. Kirk (1973) reported results generally accurate to three decimal places, with few exceptions, using his algorithm and double-precision arithmetic. TETREST involves a slight modification of Kirk's method, and does not use double-precision. Nevertheless, on a CDC CYBER 74 computer, TETREST results are still usually accurate to three decimal places. Users may evaluate for themselves the accuracy of TETREST by using, as Type III input data, joint and marginal proportions corresponding to known true values of the tetrachoric correlation. National Bureau of Standards (1959) tables are one source of such data.

The phi coefficient is calculated exactly, and is accurate to 15 significant digits on Control Data machines.

References

Bock, R.D. & Wood, R. Test theory. Annual Review of Psychology, 1971, 22, 193-219.

Carroll, J.B. The nature of the data, or how to choose a correlation coefficient. Psychometrika, 1961, 26, 347-372.

Dixon, W.J. BMD: Biomedical Computer Programs. Los Angeles: University of California Press, 1973.

Indow, T. & Samejima, F. On the results obtained by the absolute scaling model and the Lord model in the field of intelligence. Third report, the Psychology Laboratory on the Hiyoshi Campus, Keio University, Japan, 1966. (In English)

Kirk, D.B. Technique for approximating the bivariate normal correlation coefficient, ρ, and estimating tetrachoric r. Princeton, N.J. Educational Testing Service RB-71-35, June, 1971.

Kirk, D.B. On the numerical approximation of the bivariate normal (tetrachoric) correlation coefficient. Psychometrika, 1973, 38(2), 259-268.

Lord, F.M. & Novick, M.R. Statistical theories of mental test scores. Reading Massachusetts: Addison-Wesley, 1968.

McBride, J.R. & Weiss, D.J. A word knowledge item pool for adaptive ability measurement. Research Report 74-2, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974.

McNemar, Q. Psychological Statistics. New York: Wiley, 1955.

National Bureau of Standards. Tables of the bivariate normal distribution function (AMS-50). U.S. Department of Commerce, 1959.

Nie, N. Bent, D.H. & Hull, C.H. SPSS: Statistical package for the social sciences. New York: McGraw-Hill, 1970.

Appendix A

Examples of Input and Output

```
PROBLEM          NVAR=09   DATA=2     2X2=      NFMT=    KEY=1   R=BOTH
RUN NAME         DEMO--TYPE I DATA INPUT ON FILE  INPUT
INPUT FORMAT     (20X,9A1////)
OUTPUT FORMAT    (1X,20F6.2)
READ INPUT DATA

         END-OF-FILE ENCOUNTERED WHILE READING DATA
         NO.OF CASES HAS BEEN CHANGED TO       66
                   MATRIX        2
```

```
           DEMO--TYPE I DATA INPUT ON FILE  INPUT

        PRINTED BELOW IS THE JOINT FREQUENCY MATRIX,

  WITH JOINT FREQUENCY CORRECT,++,IN THE UPPER DIAGONAL,
       JOINT FREQUENCY INCORRECT ,--, IN THE LOWER DIAGONAL
       AND MARGINAL  FREQUENCY, + ,IN THE DIAGONAL
```

	1	2	3	4	5	6	7	8	9
1 *	43	22	24	17	28	23	16	31	29
2 *	21	24	14	13	17	16	14	21	20
3 *	21	30	26	10	20	13	10	19	20
4 *	4	19	14	20	14	13	13	17	16
5 *	5	13	14	17	32	22	13	22	26
6 *	19	31	26	16	15	27	13	21	22
7 *	17	25	21	12	6	27	18	17	17
8 *	19	28	24	12	7	25	20	35	26
9 *	20	30	28	14	14	29	21	25	32

LISTED BELOW IS THE MATRIX OF TETRACHORIC CORRELATIONS

VARIABLE 1	1.00								
VARIABLE 2	.69	1.00							
VARIABLE 3	.73	.45	1.00						
VARIABLE 4	-.05	.43	-.05	1.00					
VARIABLE 5	.29	.28	.49	.29	1.00				
VARIABLE 6	.56	.59	.24	.28	.63	1.00			
VARIABLE 7	.50	.69	.19	.53	-.07	.55	1.00		
VARIABLE 8	.75	.77	.50	.45	.06	.62	.77	1.00	
VARIABLE 9	.76	.76	.67	.44	.73	.77	.81	.76	1.00

LISTED BELOW IS THE MATRIX OF PHI CORRELATIONS

VARIABLE 1	1.00								
VARIABLE 2	.42	1.00							
VARIABLE 3	.46	.29	1.00						
VARIABLE 4	-.02	.28	-.03	1.00					
VARIABLE 5	.16	.18	.32	.18	1.00				
VARIABLE 6	.35	.40	.15	.18	.43	1.00			
VARIABLE 7	.27	.47	.12	.35	-.04	.35	1.00		
VARIABLE 8	.52	.52	.32	.27	.04	.41	.45	1.00	
VARIABLE 9	.52	.53	.46	.27	.51	.55	.51	.55	1.00

```
RUN NAME        DEMO---TYPE II DATA, PRINTING 2X2 TABLES AND TET MATRIX
PROBLEM         NVAR=05   DATA=3     2X2=9                                R-TET
INPUT FORMAT    (5F4.0)
OUTPUT FORMAT   (10F8.3)
OPTIONS         6
READ INPUT DATA
MATRIX          3

PRINTED BELOW ARE THE  2X2 CONTINGENCY TABLES FOR EACH VARIABLE PAIR
99.00 IS PRINTED WHEN THE PROGRAM CANNOT CONVERGE ON A SOLUTION

              -  VAR  2  +        TOTALS          PROPORTIONS
          +      48.        49.      97.     P(I,J).22395833        TETRACHORIC  R  =    .0826161   EPSILON =    .00009
VAR  1    -      52.        43.      95.     Q( I ).49479167
     TOTALS      92.       100.     192.     Q( J ).47916667                     PHI  =    .0525659
-------------------------------------------------------------------------------------------------------------------
              -  VAR  3  +        TOTALS          PROPORTIONS
          +      39.        58.      97.     P(I,J).21875000        TETRACHORIC  R  =    .0641270   EPSILON =    .00004
VAR  1    -      42.        53.      95.     Q( I ).49479167
     TOTALS     111.        81.     192.     Q( J ).42187500                     PHI  =    .0405392
-------------------------------------------------------------------------------------------------------------------
              -  VAR  4  +        TOTALS          PROPORTIONS
          +      53.        44.      97.     P(I,J).27083333        TETRACHORIC  R  =   -.1467430   EPSILON =    .00000
VAR  1    -      43.        52.      95.     Q( I ).49479167
     TOTALS      96.        96.     192.     Q( J ).50000000                     PHI  =   -.0937551
-------------------------------------------------------------------------------------------------------------------
              -  VAR  5  +        TOTALS          PROPORTIONS
          +      47.        50.      97.     P(I,J).21875000        TETRACHORIC  R  =   -.0669310   EPSILON =    .00005
VAR  1    -      42.        53.      95.     Q( I ).49479167
     TOTALS     103.        89.     192.     Q( J ).46354167                     PHI  =   -.0425418
-------------------------------------------------------------------------------------------------------------------
              -  VAR  3  +        TOTALS          PROPORTIONS
          +      42.        50.      92.     P(I,J).21875000        TETRACHORIC  R  =   -.1062507   EPSILON =    .00000
VAR  2    -      39.        61.     100.     Q( I ).47916667
     TOTALS     111.        81.     192.     Q( J ).42187500                     PHI  =   -.0672905
-------------------------------------------------------------------------------------------------------------------
              -  VAR  4  +        TOTALS          PROPORTIONS
          +      50.        42.      92.     P(I,J).21875000        TETRACHORIC  R  =   -.1307045   EPSILON =    .00000
VAR  2    -      46.        54.     100.     Q( I ).47916667
     TOTALS      96.        96.     192.     Q( J ).50000000                     PHI  =   -.0834058
-------------------------------------------------------------------------------------------------------------------
              -  VAR  5  +        TOTALS          PROPORTIONS
          +      41.        51.      92.     P(I,J).21354167        TETRACHORIC  R  =    .0541667   EPSILON =    .00003
VAR  2    -      48.        52.     100.     Q( I ).47916667
     TOTALS     103.        89.     192.     Q( J ).46354167                     PHI  =    .0344096
-------------------------------------------------------------------------------------------------------------------
              -  VAR  4  +        TOTALS          PROPORTIONS
          +      54.        57.     111.     P(I,J).20312500        TETRACHORIC  R  =    .0500502   EPSILON =    .00002
VAR  3    -      42.        39.      81.     Q( I ).42187500
     TOTALS      96.        96.     192.     Q( J ).50000000                     PHI  =    .0316386
-------------------------------------------------------------------------------------------------------------------
              -  VAR  5  +        TOTALS          PROPORTIONS
          +      53.        58.     111.     P(I,J).18750000        TETRACHORIC  R  =   -.0518551   EPSILON =    .00002
VAR  3    -      36.        45.      81.     Q( I ).42187500
     TOTALS     103.        89.     192.     Q( J ).46354167                     PHI  =   -.0327144
-------------------------------------------------------------------------------------------------------------------
              -  VAR  5  +        TOTALS          PROPORTIONS
          +   [illegible]   48.      96.     P(I,J).25000000        TETRACHORIC  R  =   -.1147666   EPSILON =    .00000
VAR  4    -   [illegible]   55.      96.     Q( I ).50000000
     TOTALS     103.        89.     192.     Q( J ).46354167                     PHI  =   -.0731113
-------------------------------------------------------------------------------------------------------------------
```

RUN NAME TYPE III DATA. EXAMPLE PROBLEMS FOR CHECKING PROGRAM, ACCURACY.
PROBLEM NVAR=10 DATA=5 NFMT=01
INPUT FORMAT (3F10.8,A1)
OPTIONS 3
READ INPUT DATA
MATRIX 1

TYPE III DATA. EXAMPLE PROBLEMS FOR CHECKING PROGRAM ACCURACY.

THE FOLLOWING TETRACHORIC CORRELATIONS WERE CALCULATED FROM THE INPUT PROPORTIONS LISTED.
CAUTION---ANY MARGINAL FREQUENCIES GREATER THAN .50 WERE CORRECTED INTERNALLY.WHICH CAUSES AN INCORRECT RESULT
NOTE ---THE SIGN OF THE CORRELATION COEFFICIENT MAY BE REVERSED IF THE JOINT PROPORTION WAS TAKEN FROM THE MINOR DIAGONAL

PROBLEM	VARIABLES		JOINT AND MARGINAL PROPORTIONS			PHI	TETRACHORIC	EPSILON
1	2,	1	.12739800	.15865525	.50000000	.26314	.50000	.00000
2	3,	1	.33333333	.50000000	.50000000	.33333	.50000	.00000
3	3,	2	.41169900	.50000000	.50000000	.64680	.85002	.00002
4	4,	1	.00134900	.15865525	.00134990	.08460	.84870	.00000
5	4,	2	.15794900	.15865525	.50000000	.43038	.89996	.00000
6	4,	3	.42821700	.50000000	.50000000	.71287	.90000	.00000
7	5,	1	.02274200	.02275013	.15865525	.35121	.95018	.00002
8	5,	2	.02275000	.02275013	.50000000	.15258	.89853	.00001
9	5,	3	.00080900	.00134990	.00134990	.59876	.95008	.00007
10	5,	4	.02275013	.15865525	.02275013	.35136	.96018	.00009
11	6,	1	.15865525	.15865525	.50000000	.43425	.96012	.00009
12	6,	2	.14500300	.15865525	.15865525	.89772	.99101	.00004
13	6,	3	.47747300	.50000000	.50000000	.90989	.99100	.00004
14	6,	4	.01971200	.02275013	.02275013	.86335	.99103	.00005
15	6,	5	.01971200	.02275013	.02275013	.86335	.99103	.00005
16	7,	1	.24203890	.50000000	.50000000	-.03184	-.05002	.00002
17	7,	2	.00000460	.02275013	.00134990	-.00477	-.19949	.00000
18	7,	3	.00000020	.00134990	.50000000	-.03675	-.74994	.00000
19	7,	4	.00011790	.02275013	.50000000	-.15100	-.75001	.00000
20	7,	5	.07178310	.50000000	.50000000	-.71287	-.90000	.00000
21	7,	6	.00000010	.02275013	.50000000	-.15258	-.90172	.00002
22	8,	1	.05054130	.50000000	.50000000	-.79783	-.95004	.00000
23	8,	2	.04516720	.50000000	.50000000	-.81933	-.96008	.00001
24	8,	3	.00000040	.15865525	.50000000	-.43425	-.95976	.00010
25	8,	4	.02252670	.50000000	.50000000	-.90989	-.99100	.00004
26	8,	5	.00000510	.15865525	.50000000	-.43422	-.95715	.K0004
27	8,	6	.03908300	.50000000	.50000000	-.84367	-.97016	.00000

LISTED BELOW IS THE MATRIX OF TETRACHORIC CORRELATIONS

```
VARIABLE    1
1.0000000

VARIABLE    2
 .4999974 1.0000000

VARIABLE    3
 .5000000  .8500160 1.0000000

VARIABLE    4
 .8486980  .8999555  .9000036 1.0000000

VARIABLE    5
 .9501760  .8985307  .9500787  .9601756 1.0000000

VARIABLE    6
 .9601208  .9910074  .9910019  .9910300  .9910300 1.0000000

VARIABLE    7
-.0500211 -.1994941 -.7499432 -.7500099 -.9000033 -.9017240 1.0000000

VARIABLE    8
-.9500391 -.9600849 -.9597555 -.9910023 -.9571481 -.9701562           1.0000000
```

Appendix B

A Complete FORTRAN Listing of PROGRAM TETREST

```
      PROGRAM TETREST(INPUT,OUTPUT,PUNCH,TAPE1,TAPE8,TAPE9,DATA,TAPE6=OU
     +TPUT,TAPE10=PUNCH,TAPE51=DATA,TAPE55=INPUT)
      COMMON/RFREQ/R(75,75)
      COMMON/FRMAT/IFMT(80),OFMT(8)
      COMMON/IOPT/OPT(8)
      COMMON/MISC/ICOL(75),NAME(8),NI,NS,NF,IR,LIU,JSPEC
      DIMENSION DUM(75)
      INTEGER OFMT
      IRUN=0
1     CALL SETUP
      CALL DATE(NOW)
      PRINT 1019,NAME,NOW
      ZL=NS
      REWIND 1
      IF(OPT(1).EQ.5.)GO TO 7
      IF(OPT(1).GE.3.)GO TO 31
      CALL FREQ1(NS,NI,IR)
      IF(OPT(1).EQ.2.)4,61
31    READ(1,IFMT)((R(I,J),J=1,NI),I=1,NI)
      IF(OPT(1).EQ.4.)4,61
4     PRINT 1003
      DO 5 I=1,NI
5     ICOL(I)=I
      PRINT 1004,(ICOL(I),I=1,NI)
      PRINT 1005
      DO 6I=1,NI
      PRINT 1006,I
      ENCODE(125,1007,DUM(1))(R(I,J),J=1,NI)
      DECODE(125,1008,DUM(1))(ICOL(J),J=1,NI)
      DO 51 J=1,NI
      IF(ICOL(J).EQ.4H    )ICOL(J)=4H   0
51     ICOL(J)=ICOL(J).A.77777777007777777777B.O.550000000000B
      IF(OPT(1).NE.4.)GO TO 6
      WRITE(9,1008)(ICOL(J),J=1,NI)
6     PRINT 1009,(ICOL(J),J=1,NI)
      ENDFILE 9
61    CALL CELLS2
      GO TO 10
7     N=0
      IF(OPT(2).GE.2.)GO TO 71
      PRINT 1012
      PRINT 1013
71    R(1,1)=1.
      DO 8 I=2,NI
      R(I,I)=1.
      K=I-1
      DO 8 J=1,K
      R(I,J)=R(J,I)=99.0
      READ (1,IFMT)PC,PI,PJ,ISIGN
      ASIGN=1.
      IF(ISIGN.NE.10H          )ASIGN=-1.
      IF(EOF(1))994,72,994
72     CALL APXTET3(PC,PI,PJ,OUT,EPS)
      IF(ASIGN.EQ.0.)ASIGN=1.
      R(J,I)=OUT*ASIGN
```

```
      PHI=(PC-(PI*PJ))/SQRT(PI*PJ*(1.-PI)*(1.-PJ))
      PHI=PHI*ASIGN
      IF(OPT(2).GE.2.)GO TO 73
      OUT=R(J,I)
      N=N+1
      PRINT 1014,N,I,J,PC,PI,PJ,PHI,OUT,EPS
73    CONTINUE
8     R(I,J)=PHI
994   IF(OPT(2).GE.2.)GO TO 10
      PRINT 1010
CCCCC.....SUPPRESS R-MATRIX
10    IF(OPT(3).EQ.6.)GO TO 1
      ISPEC=1
       IF(JSPEC.EQ.3HBOT)ISPEC=2
      IF(JSPEC.EQ.3HPHI)ISPEC=3
CCCCC.....BRANCH TO R-MATRIX OUTPUT GROUP
      IF(OPT(3).LE.4.)100,110
100   IF(OPT(3).EQ.1.)GO TO 105
      LU=6
      IF(OPT(2).EQ.3.)101,103
101   ASSIGN 102 TO NSTATE
      GO TO(201,201,202),ISPEC
      WRITE(6,9001)
      GO TO 1
102   IF(OPT(3).EQ.3.)GO TO 1
      IF(OPT(3).EQ.2.)105,110
CCCCC.....PRINT,LOWER DIAGONAL
103   ASSIGN 104 TO NSTATE
      GO TO (203,203,204),ISPEC
      WRITE(6,9001)
      GO TO 1
104   IF(OPT(3).EQ.3.)GO TO 1
      IF(OPT(3).EQ.4.)GO TO 110
CCCCC.....PUNCH SQUARE MATRICES
105   ASSIGN 108 TO NSTATE
      LU=10
      IF(OPT(2).EQ.3.)106,107
106   GO TO(201,201,202),ISPEC
      WRITE(6,9001)
107   GO TO (203,203,204),ISPEC
      WRITE(6,9001)
108   GO TO 1
CCCCC.....WRITE R-MATRICES
110   IF(OPT(2).EQ.3.)111,113
CCCCC.....SQUARE MATRICES
111   IF(ISPEC.EQ.3)GO TO 112
CCCCC.....SQUARE TET
      CALL MTRXOUT(8,1)
      IF(ISPEC.EQ.2)112,1
CCCCC.....SQUARE PHI
112   CALL MTRXOUT(8,2)
      GO TO 1
CCCCC.....LOWER DIAGONAL MATRICES
113   IF(ISPEC.EQ.3)GO TO 114
      CALL MTRXOUT(8,4)
```

```
      IF(ISPEC.EQ.2)114,1
114   CALL MTRXOUT(8,3)
      GO TO 1
CCCCC  .....SQUARE TET
201   CALL MTRXOUT(LU,1)
      IF(ISPEC.EQ.2)202,2021
CCCCC..... SQUARE PHI
202   CALL MTRXOUT(LU,2)
      GO TO 2021
203   CALL MTRXOUT(LU,4)
CCCCC..... LOWER TET
      IF(ISPEC.EQ.2)204,2021
CCCCC..... LOWER PHI
204   CALL MTRXOUT(LU,3)
2021  GO TO NSTATE,(102,104,108)
      GO TO 1
1001   FORMAT(8A10)
1002   FORMAT(1X,8A10)
1003   FORMAT(1X,*PRINTED BELOW IS THE JOINT FREQUENCY MATRIX,WITH JOINT
     2 FREQUENCY CORRECT,++,IN THE UPPER DIAGONAL,*/51X,*JOINT FREQUENCY
     3 INCORRECT ,--, IN THE LOWER DIAGONAL*,/51X,*AND MARGINAL  FREQUEN
     4CY, + ,IN THE DIAGONAL*/)
1004   FORMAT(7X,25I5)
1005   FORMAT(//)
1006   FORMAT(1X,I3,X,1H*)
1007   FORMAT(25F5.0)
1008   FORMAT(25(A4,1X))
1009  FORMAT(7X,25A5)
1010   FORMAT(1H1)
1011   FORMAT(1X,*PRINTED BELOW ARE THE  2X2  CONTINGENCY TABLES FOR EVE
     2RY ITEM PAIR*//)
1012    FORMAT(1X,*THE FOLLOWING TETRACHORIC CORRELATIONS WERE CALCULATE
     2D FROM THE INPUT PROPORTIONS LISTED.*,/1X,*CAUTION---ANY MARGINAL
     3FREQUENCIES GREATER THAN .50 WERE CORRECTED INTERNALLY,WHICH CAUSE
     4S AN INCORRECT RESULT*,/1X,*NOTE   ---THE SIGN OF THE CORRELATION
     5COEFFICIENT MAY BE REVERSED IF THE JOINT PROPORTION WAS TAKEN FROM
     6 THE MINOR DIAGONAL*,/)
1013   FORMAT(1X,7HPROBLEM,3X,9HVARIABLES,5X,30HJOINT AND MARGINAL PROPO
     2RTIONS,16X,3HPHI,7X,11HTETRACHORIC,5X,7HEPSILON/)
1014   FORMAT(1X,I5,5X,I4,*,*,I4,X,3(F10.8,5X),3(F10.5,5X))
1015   FORMAT(/1X,*VARIABLE*,I5)
1017   FORMAT(8F10.7)
1018   FORMAT(10F8.5)
1019   FORMAT(1H1,8A10,40X,A10/)
9001  FORMAT(/10(1H*),*PROGRAM ERROR---ISPEC IS OUT OF RANGE INTETREST*)
      END
```

```
      SUBROUTINE MTRXOUT(LU,ISPEC)
CCCCC.....OUTPUTS THE UPPER OR LOWER DIAGONAL OF A SQUARE MATRIX AS A SQ
CCCCC.....OR LOWER DIAGONAL MATRIX
      COMMON/RFREQ/R(75,75)
      COMMON/FRMAT/IFMT(80),OFMT(8)
```

```
      COMMON/MISC/DUM(75),NAME(8),NI,NS,NF,IR,LIU,JSPEC
      IF(LU.EQ.6)99,100
99    GO TO(105,106,106,105),ISPEC
100   GO TO(102,101,104,103),ISPEC
CCCCC.....LOWER DIAGONAL TO SQUARE
101   DO 4 I=1,NI
      IF(LU.EQ.6)1,2
1     WRITE(6,1015)I
2     IF(I.EQ.NI)GO TO 5
      K=I+1
      DO 3 J=K,NI
3     DUM(J)=R(K,I)
4     WRITE(LU,OFMT)((R(I,J),J=1,NI),(DUM(L),L=K,NI))
5     WRITE(LU,OFMT)(R(1,J),J=1,NI)
      GO TO 20
CCCCC.....UPPER DIAGONAL TO SQUARE
102   IF(LU.EQ.6)6,7
6     I=1
      WRITE(6,1015)I
7     WRITE(LU,OFMT)(R(1,J),J=1,NI)
      DO 9 I=2,NI
      IF(LU.EQ.6)71,72
71    WRITE(6,1015)I
72    IF(I.EQ.NI)GO TO 10
      DO 8 J=1,NI
8     DUM(J)=R(J,I)
      L=I+1
9     WRITE(LU,OFMT)((DUM(J),J=1,I),(R(I,K),K=L,NI))
10     WRITE(LU,OFMT)(R(I,NI),I=1,NI)
      GO TO 20
CCCCC.....UPPER DIAGONAL TO LOWER
103   DO 13 I=1,NI
      DO 11 J=1,I
11    DUM(J)=R(J,I)
      IF(LU.EQ.6)12,13
12    WRITE(6,1015)I
13    WRITE(LU,OFMT)(DUM(J),J=1,I)
      GO TO 20
1015   FORMAT(/1X,*VARIABLE*,I5)
CCCCC.....LOWER DIAGONAL TO LOWER
104   DO 15 I=1,NI
      K=I
      IF(LU.EQ.6)14,15
14    WRITE(6,1015)I
15    WRITE(LU,OFMT)(R(I,J),J=1,K)
      GO TO 20
105   WRITE(6,1051)
      GO TO 100
106   WRITE(6,1061)
      GO TO(102,101,104,103),ISPEC
20     END FILE LU
      RETURN
1051   FORMAT(*1LISTED BELOW IS THE MATRIX OF TETRACHORIC CORRELATIONS*
     +///)
1061   FORMAT(*1LISTED BELOW IS THE MATRIX OF PHI CORRELATIONS*///)
       END
```

```
      SUBROUTINE SETUP
      COMMON/RFREQ/R(75,75)
      COMMON/FRMAT/IFMT(80),OFMT(8)
      COMMON/IOPT/OPT(8)
      COMMON/MISC/ICOL(75),NAME(8),NI,NS,NF,IR,LIU,JSPEC
      DIMENSION IN(8),DUM(75)
1     NS=9999$NF=1$LIU=55$NI=0
      PRINT 2007
      CALL DATE(NOW)
      IVER=9H 1 FEB 75
      PRINT 1000,NOW,IVER
      DO 2 I=1,8
      NAME(I)=10H
      OPT(I)=I
2     OFMT(I)=10H
      OFMT(1)=10H(8F10.7)
11    DO 12 I=1,8
12    IN(I)=10H
      READ(55,2001)IN
      IF(EOF(55))991,13,991
13    PRINT 2002,IN
      IF(IN(1).EQ.10HPROBLEM   )GO TO 20
      IF(IN(1).EQ.10HRUN NAME  )GO TO 30
      IF(IN(1).EQ.10HINPUT MEDI)GO TO 40
      IF(IN(1).EQ.10HINPUT FORM)GO TO 50
      IF(IN(1).EQ.10HNO. OF CAS)GO TO 60
      IF(IN(1).EQ.10HOUTPUT FOR)GO TO 70
      IF(IN(1).EQ.10HOPTIONS   )GO TO 80
      IF(IN(1).EQ.10HMATRIX    )GO TO 90
      IF(IN(1).EQ.10HREAD INPUT)GO TO 100
      IF(IN(1).EQ.10HFINISH    )GO TO 110
      PRINT 992
      STOP2
20    DECODE(60,2003,IN(3))NI,OPT(1),OPT(4),NF,IR,JSPEC
      IF(JSPEC.EQ.3H   )JSPEC=3HTET
      IF(NI.LE.1.OR.NI.GT.75)GO TO 993
      IF(OPT(1).LT.0..OR.OPT(1).GT.5.)GO TO 993
      IF(OPT(1).EQ.0.)OPT(1)=1.
      IF(IR.EQ.5H     )IR=1H1
      IF(NF.EQ.0)NF=1
      IF(NF.LT.0.OR.NF.GT.10)GO TO 993
      IF(OPT(4).LT.0..OR.OPT(4).GT.9.)GO TO 993
      GO TO 11
30    DO 31 I=1,7
31    NAME(I)=IN(I+1)
      NAME(8)=10H
      GO TO 11
40    IF(IN(2).EQ.10HUM   CARD )41,42
41    LIU=55 $ GO TO 11
42    IF(IN(2).EQ.10HUM   TAPE )43,44
```

```
43      LIU=51 $ GO TO 11
44      IF(IN(2).EQ.10HUM   DISC )GO TO 43
        PRINT 9009
        GO TO 11
50      DO 51 I=1,80
51      IFMT(I)=10H
        DECODE(70,2004,IN(2))(IFMT(I),I=1,7)
        IF(NF.EQ.1)11,52
52      NF=7+((NF-1)*8)
        READ(55,2001)(IFMT(I),I=8,NF)
        PRINT 2002,(IFMT(I),I=8,NF)
        GO TO 11
60      DECODE(10,2005,IN(2))NS
        IF(NS.LT.1.OR.NS.GT.9999)GO TO 994
        GO TO 11
70      DECODE(70,2004,IN(2))(OFMT(I),I=1,7)
        GO TO 11
80      DECODE(10,2006,IN(2))OPT(3)
        IF(OPT(3).LE.0..OR.OPT(3).GT.6.)GO TO 995
        GO TO 11
90      DECODE(10,2006,IN(2))OPT(2)
        IF(OPT(2).LE.0..OR.OPT(2).GT.3.)GO TO 996
        RETURN
100     REWIND 1
        IF(OPT(1).EQ.5.)101,103
101     DO 102 I=2,NI
        N1=I-1
        DO 102 L=1,N1
        READ(LIU,IFMT)(DUM(J),J=1,4)
        IF(EOF(LIU))1021,102,1021
102     WRITE(1,IFMT)(DUM(J),J=1,4)
        GO TO 1022
1021     NI=I
1022     END FILE 1
        GO TO 11
103     IF(OPT(1).GE.3.)104,106
104     DO 105 I=1,NI
        READ(LIU,IFMT)(DUM(J),J=1,NI)
        IF(EOF(LIU))997,105,997
105     WRITE(1,IFMT)(DUM(J),J=1,NI)
        END FILE 1
        GO TO 11
106     DO 107 I=1,NS
        READ(LIU,IFMT)(ICOL(J),J=1,NI)
        IF(EOF(LIU))1061,107,1061
1061    NS=I-1
        PRINT 9008,NS
        GO TO 108
107     WRITE(1,IFMT)(ICOL(J),J=1,NI)
108     END FILE 1
        GO TO 11
110     STOP
991     CONTINUE
        PRINT 9001
        PRINT 9999
```

```
      STOP1
993   PRINT 9003
      PRINT 9999
      STOP3
994   PRINT 9004
      PRINT 9999
       STOP4
995   PRINT 9005
      PRINT 9999
      STOP5
996   PRINT 9006
      PRINT 9999
      STOP 6
997   PRINT 9997
      PRINT 9999
      STOP7
9001   FORMAT(/10(1H*),*ERROR---END/OF-FILE ENCOUNTERED WHILE READING TE
     2REST CONTROL CARDS*)
9003   FORMAT(/10(1H*),*ERROR---ONE OR MORE OF YOUR PROBLEM CARD PARAMET
     2ERS IS OUT OF RANGE*,/18X,*CHECK THE VALUES OF NVAR,NF,DATA AND
     32X2.*)
9004   FORMAT(/10(1H*),*ERROR---THE NO.OF CASES IS TOO LARGE OR NON-NUME
     2RIC VALUE WAS READ*)
9005   FORMAT(/10(1H*),*ERROR---THE OPTION CARD VALUE IS OUT OF RANGE OR
     2 A NON-NUMERIC VALUE WAS READ*)
9006    FORMAT(/10(1H*),*ERROR---THE MATRIX CARD VALUE IS OUT OF RANGE O
     2R A NON-NUMERIC VALUE WAS READ*)
9008   FORMAT(/10X,*END-OF-FILE ENCOUNTERED WHILE READING DATA*,/
     210X,*NO.OF CASES HAS BEEN CHANGED TO *,I8)
9009   FORMAT(/10X,*COMMENT---CHECK YOUR INPUT MEDIUM CARD FOR ERROR. IN
     2PUT MEDIUM IS ASSUMED TO BE CARDS.*)
992   FORMAT(/10(1H*),*ERROR---THE PROGRAM DID NOT READ A COMPLETE SET
     2OF CONTROL CARDS.*,/18X,*IF CONTROL CARDS ARE COMPLETE,CHECK FOR M
     3ISSPELLED DEFINITION FIELDS*)
9997   FORMAT(10(1H*),*ERROR---END-OF-FILE ENCOUNTERED WHILE READING JOI
     2NT FREQUENCY MATRIX*)
9999   FORMAT(/10X,*FURTHER DIAGNOSTICS WOULD BE MISLEADING.*)
1000   FORMAT(100X,A10/100X,*PROGRAM TETREST*,/100X,A10,2X,*REVISION*,
     +//////////)
2001   FORMAT(8A10)
2002   FORMAT(20X,8A10)
2003   FORMAT(I2,8X,F1,9X,F1,9X,I2,8X,A1,9X,A3)
2004   FORMAT(5X,A5,6A10)
2005   FORMAT(5X,I4,1X)
2006  FORMAT(5X,F1,4X)
2007  FORMAT(1H1)
      STOP
      END


      SUBROUTINE FREQ1(NS,NV,IR)
      COMMON/RFREQ/IFR(75,75)
      COMMON/FRMAT/IFMT(80)
```

```
      DIMENSION IX(75)
      REAL IFR
      NI=NV
      DO102 I=1,NI
      DO102 J=1,NI
      IFR(I,J)=0.
102   CONTINUE
      DO 107 L=1,NS
103   READ(1,IFMT)(IX(I),I=1,NI)
      DO 107 I=1,NI
      DO 107 J=I,NI
       IF(IX(I).EQ.IX(J))104,107
104   IF(IX(I).EQ.IR)105,106
105   IFR(I,J)=IFR(I,J)+1.
      GO TO 107
106   IF(I.EQ.J)GO TO 107
      IFR(J,I)=IFR(J,I)+1.
107   CONTINUE
      RETURN
      END
```

```
      SUBROUTINE CELLS2
      COMMON/RFREQ/R(75,75)
      COMMON/IOPT/OPT(8)
      COMMON/MISC/ICOL(75),NAME(8),NI,NS,NF,IR,LIU,JSPEC
      N1=NI-1
      IF(OPT(4).NE.9.)GO TO 200
      PRINT 3
3     FORMAT(*1PRINTED BELOW ARE THE  2X2 CONTINGENCY TABLES FOR EACH VA
     2RIABLE PAIR*,/1X,*99.00 IS PRINTED WHEN THE PROGRAM CANNOT CONVERG
     3E ON A SOLUTION*,//)
200   DO 227 I=1,N1
      N2=I+1
      DO 227 J=N2,NI
      A=R(I,J)
      D=R(J,I)
      E=R(I,I)
      G=R(J,J)
      B=E-A
      C=G-A
      F=C+D
      H=B+D
      ZL=A+B+C+D
      DET=(A*D)-(B*C)
      IF(DET)2002,2001,2002
2001  PHI=0.
      R(I,J)=0.
      GO TO 2003
2002  PHI=DET/SQRT(E*F*G*H)
2003  IF(JSPEC.EQ.3HPHI)GO TO 226
      IF(F.GE.E)201,202
201   ANUM=E
```

```
      IBIT=2  $  GO TO 203
202   ANUM=F
      IBIT=0
203   IF(H.GE.G)204,205
204   BNUM=G
      IBIT=IBIT+1  $  GO TO 206
205   BNUM=H
206   IF(IBIT.EQ.0)207,208
207   CELL=D $ GO TO 215
208   IF(IBIT.EQ.1)209,210
209   CELL=C  $  GO TO 215
210   IF(IBIT.EQ.2)211,212
211   CELL=B  $ GO TO 215
212   IF(IBIT.EQ.3)214,213
213    STOP20001
214   CELL=A
215   P=CELL/ZL
      QI=ANUM/ZL
      QJ=BNUM/ZL
216   IF(DET)220,226,220
220   CALL APXTET3(P,QI,QJ,R1,EPS)
      IF(DET)222,223,223
222   X=-1. $ GO TO 225
223   X=1.  $ GO TO 225
225   R(I,J)=X*ABS(R1)
      TEMP=ABS(R1)
      IF(TEMP.EQ.99.)GO TO 226
      IF(EPS.LE..0001)GO TO 226
      PRINT 2,I,J,EPS
226   R(J,I)=PHI
      IF(OPT(4).NE.9.)GO TO 227
      PRINT 1,J,B,A,E,P,R(I,J),EPS,I,D,C,F,QI,G,H,ZL,QJ,PHI
227   CONTINUE
      DO 228 I=1,NI
228   R(I,I)=1.
1     FORMAT(17X,6H-  VAR,I3,3H  +,8X,6HTOTALS,11X,11HPROPORTIONS/
     212X,1H+,3F10.0,5X,6HP(I,J),F9.8,9X,17HTETRACHORIC  R  =,F13.7,3X,
     39HEPSILON =,F10.5/1X,3HVAR,I3,5X,1H-,3F10.0,5X,6HQ( I ),F9.8/
     47X,6HTOTALS,3F10.0,5X,6HQ( J ),F9.8,20X,6HPHI  =,F13.7/1X,130(1H-)
     5)
2     FORMAT(1X,*CORRELATION(*,I3,*,*,I3,*)WAS OBTAINED USING A SUBSTAND
     2ARD CONVERGENCE VALUE OF*,F10.5)
      RETURN
      END


      SUBROUTINE APXTET3(PC,PI,PJ,RO,EPS)
      REAL M,K
      EPS=0.
      IF(PC.NE.PI)GO TO 1
      IF(PI.EQ.PJ)14,1
1     TWOPI=6.2831853072
      M=PC-PI*PJ
```

```
      IF(PI.GT..5) PI=1.-PI
      IF(PJ.GT..5) PJ=1.-PJ
      QI=1.-PI
      QJ=1.-PJ
      ZI=CDFNI(QI)
      ZJ=CDFNI(QJ)
      H=EXP(.5*(-1.*(ZI**2)))/SQRT(TWOPI)
      K=EXP(.5*(-1.*(ZJ**2)))/SQRT(TWOPI)
      R=M/(H*K)
      R=R-(ZI*ZJ)*(R**2/2.)
      IF(R.GT..80)4,2
2     IF(R.LT.-.80)3,5
3     R=-.80 $ GO TO 5
4     R=.80
5     RUT=ZI**2 + ZJ**2
      J=1
100   DO 11 I=1,50
      IF(R.GE.1.)8,6
6     IF(R.LE.-1.)7,9
7     R=-.97 $ GO TO 9
8     R=.97
9     CONTINUE
      DIV=1.-R**2
      FPRI=EXP((-1.*(RUT-2.*ZI*ZJ*R))/(2.*DIV))/(TWOPI*SQRT(DIV))
      FOFR=GAUSS8(R,ZI,ZJ)
      R1=R-(FOFR-M)/FPRI
      EPS=ABS(R1-R)
      IF(EPS.LT..0001)15,10
10     R=R1
11     CONTINUE
      J=J+1
      IF(J.GT.2)13,12
12    R=.55 $ GO TO 100
13    IF(EPS.LE..001)GO TO 15
      R=99.  $ GO TO 15
14    R=1.
15    RO=R
      RETURN
      END
      FUNCTION GAUSS8(R,ZI,ZJ)
      DIMENSION U(8),W(8),U2(8)
      DATA(U(I),I=1,8)/.0198550718,.1016667613,.2372337950,.4082826788,
     +.5917173212,.7627662050,.8983332387,.9801449282/
      DATA(W(I),I=1,8)/.0506142681,.1111905172,.1568533229,.1813418917,
     +.1813418917,.1568533229,.1111905172,.0506142681/
      DATA(U2(I),I=1,8)/.0003942239,.0103361304,.0562798735,
     +.1666947458,.3501293883,.5818122834,.8070026078,.9606840804/
      SUM=0.
      TEM1=R/6.2831853072
      TEM2=2.*ZI*ZJ*R
      R2=R**2
      RUT=ZI**2+ZJ**2
      DO 10 I=1,8
      ADIV=1.-R2*U2(I)
      G=EXP((-1.*(RUT-TEM2*U(I)))/(2.*ADIV))/SQRT(ADIV)
```

```
10      SUM=SUM+G*W(I)
        GAUSS8=SUM*TEM1
        RETURN
        END
```

DISTRIBUTION LIST

Navy

4 Dr. Marshall J. Farr, Director
Personnel and Training Research Programs
Office of Naval Research (Code 458)
Arlington, VA 22217

1 ONR Branch Office
495 Summer Street
Boston, MA 02210
ATTN: J. Lester

1 ONR Branch Office
1030 East Green Street
Pasadena, CA 91101
ATTN: E.E. Gloye

1 ONR Branch Office
536 South Clark Street
Chicago, IL 60605
ATTN: M.A. Bertin

6 Director
Naval Research Laboratory
Code 2627
Washington, DC 20390

12 Defense Documentation Center
Cameron Station, Building 5
5010 Duke Street
Alexandria, VA 22314

1 Special Assistant for Manpower
OASN (M&RA)
Pentagon, Room 4E794
Washington, DC 20350

1 LCDR Charles J. Theisen, Jr., MSC, US
4024
Naval Air Development Center
Warminster, PA 18974

1 Chief of Naval Reserve
Code 3055
New Orleans, LA 70146

1 Dr. Lee Miller
Naval Air Systems Command
AIR-413E
Washington, DC 20361

1 CAPT John F. Riley, USN
Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Chief
Bureau of Medicine & Surgery
Research Division (Code 713)
Washington, DC 20372

1 Chairman
Behavioral Science Department
Naval Command & Management Division
U.S. Naval Academy
Luce Hall
Annapolis, MD 21402

1 Chief of Naval Education & Training
Naval Air Station
Pensacola, FL 32508
ATTN: CAPT Bruce Stone, USN

1 Mr. Arnold Rubinstein
Naval Material Command (NAVMAT 03424)
Room 820, Crystal Plaza #6
Washington, DC 20360

1 Commanding Officer
Naval Medical Neuropsychiatric
Research Unit
San Diego, CA 92152

1 Director, Navy Occupational Task
Analysis Program (NOTAP)
Navy Personnel Program Support
Activity
Building 1304, Bolling AFB
Washington, DC 20336

1 Dr. Richard J. Niehaus
Office of Civilian Manpower Management
Code 06A
Washington, DC 20390

1 Department of the Navy
Office of Civilian Manpower Management
Code 263
Washington, DC 20390

1 Chief of Naval Operations (OP-987E)
Department of the Navy
Washington, DC 20350

1 Superintendent
Naval Postgraduate School
Monterey, CA 93940
ATTN: Library (Code 2124)

1 Commander, Navy Recruiting Command
4015 Wilson Boulevard
Arlington, VA 22203
ATTN: Code 015

1 Mr. George N. Graine
Naval Ship Systems Command
SHIPS 047C12
Washington, DC 20362

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 3054
ATTN: Dr. Norman J. Kerr

1 Dr. William L. Maloy
Principal Civilian Advisor
for Education & Training
Naval Training Command, Code 01A
Pensacola, FL 32508

1 Dr. Alfred F. Smode, Staff Consultant
Training Analysis & Evaluation Group
Naval Training Equipment Center
Code N-00T
Orlando, FL 32813

1 Dr. Hanns H. Wolff
Technical Director (Code N-2)
Naval Training Equipment Center
Orlando, FL 32813

1 Chief of Naval Training Support
Code N-21
Building 45
Naval Air Station
Pensacola, FL 32508

1 Dr B. Rimland
Navy Personnel R&D Center
San Diego, CA 92152

5 Navy Personnel R&D Center
San Diego, CA 92152
ATTN: Code 10

Army

1 Headquarters
U.S. Army Administration Center
Personnel Administration Combat
Development Activity
ATCP-HRO
Ft. Benjamin Harrison, IN 46249

1 Armed Forces Staff College
Norfolk, VA 23511
ATTN: Library

1 Commandant
United States Army Infantry School
ATTN: ATSH-DET
Fort Benning, GA 31905

1 Deputy Commander
U.S. Army Institute of Administration
Fort Benjamin Harrison, IN 46216
ATTN: EA

1 Dr. Stanley L. Cohen
U.S. Army Research Institute
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Ralph Dusek
U.S. Army Research Institute
1300 Wilson Boulevard
Arlington, VA 22209

1 Mr. Edmund F. Fuchs
U.S. Army Research Institute
1300 Wilson Boulevard
Arlington, VA

1 Dr. J.E. Uhlaner, Technical Director
U.S. Army Research Institute
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Joseph Ward
U.S. Army Research Institute
1300 Wilson Boulevard
Arlington, VA 22209

1 HQ USAREUR & 7th Army
ODCSOPS
USAREUR Director of GED
APO New York 09403

Air Force

1 Research Branch
AF/DPMYAR
Randolph AFB, TX 78148

1 Dr. G.A. Eckstrand (AFHRL/AS)
Wright-Patterson AFB
Ohio 45433

1 Dr. Ross L. Morgan (AFHRL/AST)
Wright-Patterson AFB
Ohio 45433

1 AFHRL/DOJN
Stop #63
Lackland AFB, TX 78236

1 Dr. Robert A. Bottenberg (AFHRL/SM)
Stop #63
Lackland AFB, TX 78236

1 Dr. Martin Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Major P.J. DeLeo
Instructional Technology Branch
AF Human Resources Laboratory
Lowry AFB, CO 80230

1 AFOSR/NL
1400 Wilson Boulevard
Arlington, VA 22209

1 Commandant
USAF School of Aerospace Medicine
Aeromedical Library (SUL-4)
Brooks AFB, TX 78235

Marine Corps

1 Mr. E.A. Dover
Manpower Measurement Unit (Code MPI)
Arlington Annex, Room 2413
Arlington, VA 20380

1 Commandant of the Marine Corps
Headquarters, U.S. Marine Corps
Code MPI-20
Washington, DC 20380

1 Director, Office of Manpower Utiliz...
Headquarters, Marine Corps (Code MP...
MCB (Building 2009)
Quantico, VA 22134

1 Dr. A.L. Slafkosky
Scientific Advisor (Code RD-1)
Headquarters, U.S. Marine Corps
Washington, DC 20380

Coast Guard

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (G-P-...
U.S. Coast Guard Headquarters
Washington, DC 20590 (All)

1 Lt. Col. Henry L. Taylor, USAF
Military Assistant for Human Resources
OAD (E&LS) ODDR&E
Pentagon, Room 3D129
Washington, DC 20301

1 Col. Austin W. Kibler
Advanced Research Projects Agency
Human Resources Research Office
1400 Wilson Boulevard
Arlington, VA 22209

Other Government

1 Dr. Lorraine D. Eyde
Personnel Research and Development
Center
U.S. Civil Service Commission
1900 E. Street, N.W.
Washington, DC 20415

1 Dr. William Gorham, Director
Personnel Research and Development
Center
U.S. Civil Service Commission
1900 E. Street, N.W.
Washington, DC 20415

1 Dr. Vern Urry
Personnel Research and Development
Center
U.S. Civil Service Commission
1900 E. Street, N.W.
Washington, DC 20415

1 Dr. Eric McWilliams, Program Manager
Technology and Systems, TIE
National Science Foundation
Washington, DC 20550

1 Dr. Andrew R. Molnar
Technological Innovations in Educat...
National Science Foundation
Washington, DC 20550

1 U.S. Civil Service Commission
Federal Office Bldg.
Chicago Regional Staff Div.
Attn: C. S. Winiewicz
Regional Psychologist
230 So. Dearborn St.
Chicago, IL 60604

Miscellaneous

1 Dr. Scarvia B. Anderson
Educational Testing Service
17 Executive Park Drive, N.E.
Atlanta, GA 30329

1 Dr. John Annett
The Open University
Milton Keynes
Buckinghamshire
ENGLAND

1 Dr. Richard C. Atkinson
Stanford University
Department of Psychology
Stanford, CA 94305

1 Dr. Gerald V. Barrett
University of Akron
Department of Psychology
Akron, OH 44325

1 Dr. Bernard M. Bass
University of Rochester
Management Research Center
Rochester, NY 14627

1 Mr. Kenneth M. Bromberg
Manager - Washington Operations
Information Concepts, Inc.
1701 North Fort Myer Drive
Arlington, VA 22209

1 Dr. Kenneth E. Clark
University of Rochester
College of Arts & Sciences
River Campus Station
Rochester, NY 14627

1 Dr. Ronald P. Carver
School of Education
University of Missouri - Kansas City
Kansas City, Missouri 64110

1 Century Research Corporation
4113 Lee Highway
Arlington, VA 22207

1 Dr. Rene' V. Dawis
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

1 Dr. Norman R. Dixon
Room 170
190 Lothrop Street
Pittsburgh, PA 15260

1 Dr. Robert Dubin
University of California
Graduate School of Administration
Irvine, CA 92664

1 Dr. Marvin D. Dunnette
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

1 ERIC
Processing and Reference Facility
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Victor Fields
Montgomery College
Department of Psychology
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
American Institutes for Research
Foxhall Square
3301 New Mexico Avenue, N.W.
Washington, DC 20016

1 Dr. Robert Glaser, Director
University of Pittsburgh
Learning Research & Development Center
Pittsburgh, PA 15213

1 Mr. Harry H. Harman
Educational Testing Service
Princeton, NJ 08540

1 Dr. Richard S. Hatch
Decision Systems Associates, Inc.
11428 Rockville Pike
Rockville, MD 20852

1 Dr. M.D. Havron
Human Sciences Research, Inc.
7710 Old Spring House Road
West Gate Industrial Park
McLean, VA 22101

1 HumRRO
Division No. 3
P.O. Box 5787
Presidio of Monterey, CA 93940

1 HumRRO
Division No. 6, Library
P.O. Box 428
Fort Rucker, IL 36360

1 Dr. Lawrence B. Johnson
Lawrence Johnson & Associates, Inc.
200 S. Street, N.W., Suite 502
Washington, DC 20009

1 Dr. Milton S. Katz
MITRE Corporation
Westgate Research Center
McLean, VA 22101

1 Dr. Steven W. Keele
University of Oregon
Department of Psychology
Eugene, OR 97403

1 Dr. David Klahr
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Ernest J. McCormick
Purdue University
Department of Psychological Sciences
Lafayette, IN 47907

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Park
Goleta, CA 93017

1 Mr. Edmond Marks
405 Old Main
Pennsylvania State University
University Park, PA 16802

1 Dr. Leo Munday, Vice-President
American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Dr. Joseph W. Rigney
University of Southern California
Behavioral Technology Laboratories
3717 South Grand
Los Angeles, CA 90007

1 Dr. Leonard L. Rosenbaum, Chairman
Montgomery College
Department of Psychology
Rockville, MD 20850

1 Dr. George E. Rowland
Rowland and Company, Inc.
P.O. Box 61
Haddonfield, NJ 08033

1 Dr. Arthur I. Siegel
Applied Psychological Services
404 East Lancaster Avenue
Wayne, PA 19087

1 Dr. C. Harold Stone
1428 Virginia Avenue
Glendale, CA 91202

1 Mr. Dennis J. Sullivan
725 Benson Way
Thousand Oaks, CA 91360

1 Dr. Benton J. Underwood
Northwestern University
Department of Psychology
Evanston, IL 60201

1 Carl R. Vest, Ph.D.
Battelle
Memorial Institute
Washington Operations
2030 - M Street, N.W.
Washington, D.C. 20036

1 Dr. Roger A. Kaufman
United States International Univ
Graduate School of Leadership an
Human Behavior
Elliott Campus
8655 E. Pomerada Road
San Diego, CA 92124

1 Dr. Anita West
Denver Research Institute
University of Denver
Denver, CO 80210

Previous Reports in this Series

73-1. Weiss, D.J. & Betz, N.E. Ability Measurement: Conventional or Adaptive? February 1973 (AD 757788).

73-2. Bejar, I.I. & Weiss, D.J. Comparison of Four Empirical Differential Item Scoring Procedures. August 1973.

73-3. Weiss, D.J. The Stratified Adaptive Computerized Ability Test. September 1973 (AD 768376).

73-4. Betz, N.E. & Weiss, D.J. An Empirical Study of Computer-Administered Two-stage Ability Testing. October 1973 (AD 768993).

74-1. DeWitt L.J. & Weiss, D.J. A Computer Software System for Adaptive Ability Measurement. January 1974 (AD 773961).

74-2. McBride, J.R. & Weiss, D.J. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974 (AD 781894).

74-3. Larkin, K.C. & Weiss, D.J. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974 (AD 783553).

74-4. Betz, N.E. & Weiss, D.J. Simulation Studies of Two-stage Ability Testing. October 1974 (AD A001230)

74-5. Weiss, D.J. Strategies of Adaptive Ability Measurement. December 1974. (AD A004270)

75-1. Larkin, K.C. & Weiss, D.J. An Empirical Comparison of Two-stage and Pyramidal Adaptive Ability Testing. February 1975.

AD Numbers are those assigned by the Defense Documentation Center, for retrieval through the National Technical Information Service

---

Copies of these reports are available, while supplies last, from:

Psychometric Methods Program
Department of Psychology
University of Minnesota
Minneapolis, Minnesota 55455